On the possibility of opening the digital materials of the National Library of Estonia for text and data mining

Summary of legal analysis

RaRa NATIONAL LIBRARY OF ESTONIA

open digital libraries

Creative Europe MEDIA

Background information

About the project

- The National Library of Estonia (hereinafter NLE), together with the National Library of the Netherlands and the Austrian National Library is participating in the project ‘Open Digital Libraries for Creative Users’ (2021-2023).
- The overall aim of the project is, first, to develop and test new creative methods for opening up the partners' digital collections to current and potential users and, second, to increase the skills and knowledge of the library sector on how to enable wider access to digital collections and thus their creative use.
- The NLE's role in the project involves the creation of a virtual lab. The virtual lab entails a web environment and a service that supports it, which would provide a convenient and relevant opportunity for potential interested parties to use the data from digital collections for research and innovation purposes.

About the legal analysis

- April 2022 saw the completion of the legal analysis ‘The opportunities, limitations and risks of the National Library in opening its digital materials for text and data mining: a perspective regarding intellectual property and personal data protection’ [„Rahvusraamatukogu võimalused, piirangud ja riskid oma digitaalse ainese teksti- ja andmekaeveks avamisel: intellektuaalse omandi ja isikuandmete kaitse perspektiiv“]. It was prepared by Professor Aleksei Kelli.
- The purpose of the analysis was to gain insight into the extent to which and how the NLE would be able to open up all of its digital materials for text and data mining. The analysis covered both material protected and unprotected under copyright and was intended to provide solutions on how to make as much digital material as possible easy to process for text and data mining purposes.

A summary of the legal analysis was prepared by Marianne Meiorg.

The summary does not deal with the use of copyright-free materials for text and data mining purposes, as the use of such materials is not restricted.

Summary

The use of digital material for text and data mining is limited by copyright and personal data protection regulation.

The problem with restrictions arising from copyright is the lack of legal clarity:

- there is no established practice of application in either Estonia or the EU;
- the adopted regulation is unclear and patchy (important issues have not been fully regulated);
- legal clarity is expected to come only after the European Court of Justice has issued its findings on this matter.

Copyrighted material may be used for text and data mining:

- for scientific and research purposes – the rightsholder cannot prohibit such use;
- for other purposes – the rightsholder can prohibit such use and claim remuneration for the use of their works.

Personal data protection regulation does not prohibit the use of digital materials containing personal data for text and data mining. This can be done both within the framework of research and within the framework of various types of self-expression (journalistic, artistic, academic or literary).

Text and data mining for science and research purposes

Who can participate?

- research institutions – universities, a university and private sector partnership is also possible;
- cultural heritage institutions, including libraries.

What is permitted?

The rightsholder cannot prohibit text and data mining for science and research purposes. It should be noted that the NLE **may not grant access** to copyrighted material **outside its premises**. Thus, the material cannot be granted remote virtual access, for example, through an environment that requires logging in.

Researchers **have the right** to **make copies of the digital material** for text and data mining purposes. In doing so, it must be based on the principle that such activities would 'not unreasonably prejudice the legitimate interests of the author'. Thus, the volume of digital material to be copied must be clearly justified and defined in the description of the specific purpose of the text and data mining. Secure storage must be ensured and, among other things, handover of the documents material and the process must be documented. The NLE should consider the digital marking of the copies to avoid disputes and simplify verification.

How to make it happen?

Proposal:

1. Conclude a framework agreement between the NLE and the research institution – content: secure storage of copies, use of copies only for text and data mining and the liability of the research institution for possible damage caused by the researcher.

2. Under the framework agreement, the researcher submits to the NLE an application together with a research plan explaining the purpose of the study and the scope/volume of digital material required for it.

3. **The NLE must evaluate each project individually.**

- It is evaluated in what volume and what exactly the researcher needs to achieve the goal:
  - from the copyright perspective – the proportionality of the volume of the requested material is evaluated from the point of view of the copyrights holder, i.e. so that the author's rights are minimally damaged;
  - from the perspective of the protection of personal data – evaluation of the potential damage that opening up material containing personal data for text and data mining could cause to data subjects; in particular, if the data subject has previously restricted access to the work.
- The evaluation is carried out on the basis of the research plan submitted by the researcher to the NLE.
- Based on the research plan and its decision, the NLE will create a package of digital materials that will be made available to the researcher.
  - The copying of parts of works is preferable to the copying of entire works.

4. Conclusion of an agreement between the NLE and the researcher – the researcher agrees to the accompanying obligations and the NLE agrees to the obligation to provide the researcher with the necessary volume of digital material to use.

Text and data mining for other purposes

Who can participate?
Whoever, including private companies.

What is permitted?

- The rightsholder may prohibit the use of their works for purposes not related to research and science.
- Copies may be made and retained for as long as necessary for the purposes of text and data mining.
- **The NLE must evaluate each project individually.**
- Conclusion of an agreement between the NLE and the interested party – the interested party agrees to the accompanying obligations and the NLE agrees to the obligation to provide them with the necessary volume of digital material to use.

In such a case, the NLE may provide text and data mining services for a fee. In this case, the restrictions on text and data mining for non-scientific purposes do not apply, as the NLE as a cultural heritage institution cannot be restricted in enabling text and data mining. Only the result of the mining is handed over, so that the copyrights will not be damaged.

The goal of the project is to create simple, modern and versatile opportunities for researchers, students, lecturers and other interested parties to use and work with the digital resources of the National Library of Estonia (NLE), specifically for text and data mining.

The project partnerts – the Royal Library of the Netherlands (Koninklijke Bibliotheek) and the Austrian National Library (Österreichischen Nationalbibliothek) – gave valuable input during the creation of these opportunities through the development of the virtual lab service.

The project was financed through the "Creative Europe" programme of the European Union. The content presented in this report reflects only the views of the authors. The European Education and Culture Executive Agency and the European Commission are not responsible for the use of the information contained therein.

RaRa NATIONAL LIBRARY OF ESTONIA

open digital libraries

Creative Europe MEDIA